IJETCR Open Access

ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)

Available Online at www.ijetcr.org

Volume 7; Issue 5; September-October 2019; Page No. 01-09

MODEL ARABIC SENTIMENT ANALYSIS USING DIFFERENT MACHINE LEARNING ALGORITHMS IN HOTELS DOMAINS

Yousra F. Gad

Sudan University of Science and Technology, Collage of Computer Science and Information Technology

E-mail: yousragad221@gmail.com

Conflicts of Interest: Nil

Corresponding author: Yousra F. Gad

ABSTRACT

With the popularity of Social Medias, sentiment analysis turns into the most important technique for understanding the opinion of users toward products or services. Arabic users are writing their comments using unstructured non-grammatical colloquial Arabic language, which made a complex challenge for sentiment analysis to operate without doing a lot of cleaning preprocessing stage. In this respect, this paper proposes a model on Arabic sentiment analysis with different classification algorithms using manual Arabic lexicon in hotels domain. Our experiment shows Logistic Regression and Support Vector Machine has the highest accuracy with 89% than other classifiers.

Keywords: Sentiment analysis, Arabic language, Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Recurrent Neural Networks (RNN)

1. Introduction

Online reviews play an important role in determining the success of electronic businesses. Nearly today, people usually use social networks to express their opining about products or services. These opinions are written in unstructured textual form and contain a lot of noises that need to process. Therefore, sentiment analysis is the right method to extract valuable information and organizes it in a way that helps to understand customer's needs. Sentiment analysis plays a very important role in businesses and organizations with a huge amount of documents on the internet that contain user's reviews about products, items, services and etc. It also helps in reducing the time and money to determine their customers' opinions about its products and competitors' products using a survey or other techniques. Sentiment analysis is done with many languages including English, Spanish, French, Chinese, Arabic language and etc. [5]. This paper is concerned with studying sentiment analysis for Arabic hotel's reviews using different classification models. Arabic sentiment analysis is a challenging task; because it comes in many forms [1]. In the Middle East, Arab people usually speak Arabic with colloquial, not in Modern Standard Arabic (MSA). The Arabic language is difficult to handle because of the various forms of writing different words such as (Alef (أ, (Yaa Magsoura (ى ((and needs lots of preprocessing to do the analysis. Another problem with Arabic is conflict a noun with positive polarities such as Hakeem "حكيم" which means wise (adjective) in English language and it can be a name of the person or description of a doctor in the Arabic language [3,32]. In addition, most of the researches that develop Arabic lexicons for sentiment analysis are not freely available online. The rest of this paper is structured as follows: Section II, we give light on sentiment analysis. In Section III, we discuss some related works, while in section IV presents the proposed model with the extracted dataset and software that used for the experiment. Many supervised classifications with a discussion of experimentation results in Section V. Finally, in Section VI we conclude of this paper.

2. Sentiment Analysis

Sentiment Analysis, also known as Opinion Mining, is part of Natural Language Processing (NLP) process that analysis the feelings in textual reviews and retrieve information about user reaction for particular product or services [4]. Usually, it determines the opinion of the user by assigning text values to the opinions into positive, negative or neutral. Furthermore, sentiment analysis is used to extract features for products or services when there is no rating is available, and also helps the user with the process of decision making, whether to accept that product or rejects it; based on the pros and cons of that product. [6, 7, 8, 9, 10, 11, 12] Those opinions are written in a textual format and contains a lot of noises; which makes the sentiment analysis task very difficult [13, 14]. There are a lot of tools that use a pre-processing step to clean up the data. Pre-processing involves many steps include such as POS tagging, Stop-word removal, Stemming, and Lemmatization and etc. Many famous tools that deal with unstructured data such as NLTK package, Stanford's Core NLP Suite, Natural Language Toolkit, Apache OpenNLP and etc. Sentiment analysis is classified into two categories: Machine learning approach and Lexicon based approach. The machine learning approach is divide into supervised machine learning which learns a classification model on labeled data, and unsupervised approaches which classify unlabeled data [15,16, 17]. Lexicon based approach get the polarity of the word (or the phrases) using a dictionary-based approach or corpus-based approach [18]. There are many typical algorithms are famous for classification such as Decision trees, Rule-based induction, Neural networks, Support Vector Machine (SVM), and Bayesian networks and etc. [19, 20]. Studies show 50% of users searching depends on other people opinion and recommendations before they buy any products. They tend to look for other user's perspective opinion (reviews, the rating given) describing how was their experiments with that product and considering it as being more trusted compared to watching only product advertising [21]. Sentiment analysis also helps commercial companies for improving their performance and present better products [22, 34]. User's reviews also contain a description of products or services (e.g. the meal was delicious, the camera was over-priced), which used to identify the features in the opinions [23].

3. Related Works

This paper is concerned with the Arabic language; many challenges are been discussed according to limited of research compared to another language such as English and European languages [24,31]. Most Arabic countries spoke their language by colloquial, not in a Modern Standard Arabic language (MSA) [5,33]. The Arabic language can be divided into two parts Classical and Modern. Classical Arabic is the original language spoken by the Arabs previously and considers the Quran language, while the MSA is most widely used in Arab countries since the 19th century.

In [25] proposed a Tool called colloquial Non-Standard Arabic-Modern Standard Arabic-Sentiment Analysis (CNSA-MSA-SAT). they built 18 polarities manually consist of 1,080 Arabic reviews from more than 60 different social media and news websites to calculating the polarities of colloquial Arabic and MSA reviews. Their domains are books, places, technology, education, society, movies, products, and politics. Experimental results showed high accuracy with 90% of the proposed tool using K-Nearest Neighbor (KNN) classifier.

Authors in [26] discussed the impact of sentiment analysis in health services. It detailed the process of collecting Arabic tweets, preprocessing stage in Arabic text including removing irrelevant data and normalizing the text. Their experiments were processed by several Machine Learning algorithms including NB, LR, SVM, DNNs, and CNNs. The results showed that SVM has the highest accuracy with 91%.

Shoukry et al [27] study the effect of Arabic sentence-level to perform sentiment analysis on tweets. Features are extracted using unigram and bigram. Multi classifiers are used and the experiment showed that SVM has the highest accuracy with 72% among other classifiers.

Mustafa Hammad et al. [28] proposed an analysis of Arabic Reviews. Their data set is collected from Jordanian hotels' customers' reviews and the combination of Arabic reviews and comments from Facebook, Twitter, and YouTube. They categorized the data into of the three polarities; positive, negative, and neutral. Their experiments for applying sentiment analysis are based on many Machine Learning algorithms (SVM, BPNN, Naïve Bayes, and Decision Tree) using Rapid Miner tools. Their results show that SVM is the most accurate machine learning algorithm and obtain good accuracy. However, they mentioned that the data is very limited with memory problem in Rapid Miner tools and need to be extended.

Bashar Al Shboul et al. [29] focused on Arabic sentiment analysis and the effect of the preprocessing stage for classification. They collect recent dataset made available publicly for the Arabic language called the Large Arabic Book Reviews (LABR) dataset, which consists of Arabic reviews based on a 5-star rating. They performed preprocessing by dropping non-Arabic reviews and cleaning up unwanted content. They spilled the data into an unbalanced dataset and the other on the smaller balanced dataset. Many Machine Learning algorithms are used include Decision Tree (J48), Decision Table, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naive Bayes (NB) and Multinomial NB (MNB), and Voting which combines different classifiers using rule (average of probabilities). The results showed that MNB had the best performance for the two datasets, balanced and unbalanced, with slightly higher accuracy for the smaller balanced dataset. ElSahar et al [30] build large multi-domain datasets for Sentiment Analysis in both Modern Standard Arabic and/Dialect. The datasets were scrapped from multiple domains including movies, hotels, restaurants, and products. Unigrams and bigrams were extracted from the dataset. The experimental results show that the top-performing classifier was SVM with 82% accuracy.

4. Proposed Model

In this section, we give a description of our proposed model for analyzing the sentiments in the hotel's reviews. Fig.1 we show the components of our model that consists of three layers. The first layer is a concern with cleaning the extracted dataset in the pre-processing stage. The second layer is the calculated polarity of each review and labels it with either positive or negative using Arabic manual lexicon. Finally, in the last layer, we applied different classifiers and calculated their accuracy performance.

A. Dataset

The dataset extracted from Tripadvisor contains 32,400 hotels reviews from different regions including Sudan, Dubai, Egypt, Morocco, Yemen, Qatar, Saudi Arabia, and Lebanon in time between 2017 and 2018. The data collected is in a textual form which contains a lot of noises and not cleans. The pre-processing stage is necessary to filter those noises and convert it into a suitable form for pattern discovery.

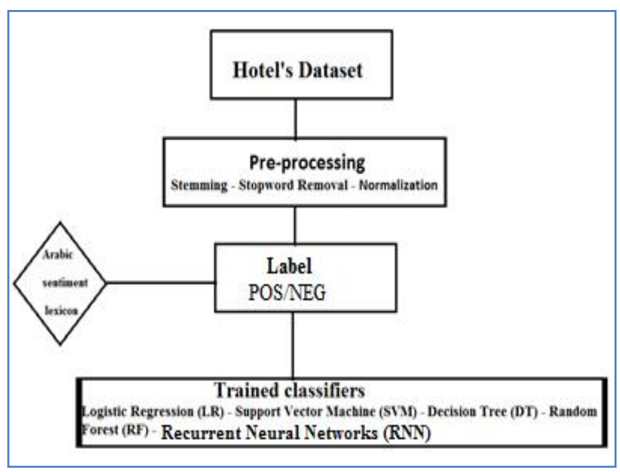

Figure 1: Proposed Arabic Sentiment Model

Pre-processing is divided into 3 steps:

1. Stemming

Stemming is a process that uses to return the word to its origin. For example, the word “يؤمنون” came from the root word “يؤمن”.

2. Remove stop-words

Stop-words are a list of words that have no meaning for NLP research. For example, (من, على, في) can be found in a sentence (الجو جميل في الفندق).

3. Normalization

The normalization process is used to organize the words to a suitable form. For example the word "راااااااائع" to "رائع".

B. Arabic sentiment lexicon

We create an Arabic sentiment lexicon which contains an amount of 2000 words extracted from the collected dataset and assign a score either positive one (+1) for positive words or a negative one(-1) for negative words. To calculating the sentiment strength we sum up all words that matched in the lexicon and give an overall score. The final score helps to indicate the strength of sentiment for each review and assign a label if it is a positive class (1) or negative class (0).

C. Trained Classifiers

We used several Machine Learning (ML) techniques including [36, 37]:

1. **Logistic Regression (LR):** Logistic model or sometimes called the logit model is a statistical method that analyzes data to the model binary dependent variable. Logistic regression estimates the probability of $P(y|x)$ by extracting a set of features from the input with weights and combining them linearly.

2. **Support Vector Machines (SVM)**: SVM is one of the supervised learning models that analyze data for classification purpose.

3. **Decision Tree (DT):** DT represents decision making and uses a tree-like model for regression and classification problems.

4. **Random Forests (RF):** Random Forests uses many classification trees, where each tree gives a vote for class. The forest calculates the most voted classification.

5. **Recurrent Neural Networks (RNN):** RNNs are a type of Neural Network that is different from traditional neural networks that each input and outputs are independent of each other; the RNNs used the output from the previous step as input to the current step.

5. **Experimental & Discussion**

Our experiments were performed on Intel corei7 CPU, 16GB of main memory, and Windows 7 Ultimate. The programing language used is Python version 3.5.0, Jupyter Notebook version 4.4.0 and NLTK for natural language processing, and Keras package for RNN deep learning classification.

In Pre-processing stage we used the following tools:

1. **Stemming**: For the stemming process, we use a light Arabic Stemming algorithm called ARLSTem stemmer [35]. ARLSTem stemmer is non-dictionary base and works by removing the affixes from the word.

2. **Stop-word Removal**: As for stop-words remove we use python Ntlk package for Natural Language Processing (NLP) and then we create a file with all Arabic common stop-words list and plug it to Ntlk files.

3. **Normalization**: we create a function that removes the repeated letter and use punctuation function.

We apply fold cross-validation method where the dataset is partitioned into 10 k different folds for training and testing. To evaluate the performance of our proposed model, we used precision-recall, f1-score, accuracy, and receiver operating characteristic (ROC) measurements. Our proposed model performed reasonably well as shown in fig 2 and fig 3. The recall rates in all classifiers are shows a lower score because of our lexicons lack coverage of all the Arabic sentiment words and even some words could be written in different form such as (I like it) "اعجبني" and "عجبني" where the manual lexicon only contains one of them. After calculating the accuracy, the classification results indicate that both Logistic Regression and Support Vector Machine have the highest accuracy with 89%.

Table 3: Classification Results

Machine Learning Techniques	Accuracy
Logistic Regression	**89.0%**
Support Vector Machine	**89.0%**
Decision Tree	85.0%
Random Forest	80.0%
Recurrent Neural Networks	79.0%

Logistic Regression

Classification Report

	precision	recall	f1-score	support
0	0.79	0.60	0.68	166
1	0.90	0.96	0.93	644
micro avg	0.89	0.89	0.89	810
macro avg	0.85	0.78	0.81	810
weighted avg	0.88	0.89	0.88	810

Support Vector Machine

Classification Report

	precision	recall	f1-score	support
0	0.77	0.65	0.70	166
1	0.91	0.95	0.93	644
micro avg	0.89	0.89	0.89	810
macro avg	0.84	0.80	0.82	810
weighted avg	0.88	0.89	0.88	810

Random Forest

Classification Report

	precision	recall	f1-score	support
0	0.00	0.00	0.00	166
1	0.80	1.00	0.89	644
micro avg	0.80	0.80	0.80	810
macro avg	0.40	0.50	0.44	810
weighted avg	0.63	0.80	0.70	810

Recurrent Neural Network

Classification Report

	precision	recall	f1-score	support
0	0.00	0.00	0.00	68
1	0.79	1.00	0.88	256
micro avg	0.79	0.79	0.79	324
macro avg	0.40	0.50	0.44	324
weighted avg	0.62	0.79	0.70	324

Decision Tree

Classification Report

	precision	recall	f1-score	support
0	0.72	0.48	0.57	166
1	0.88	0.95	0.91	644
micro avg	0.85	0.85	0.85	810
macro avg	0.80	0.71	0.74	810
weighted avg	0.84	0.85	0.84	810

Figure 2: Classification Reports

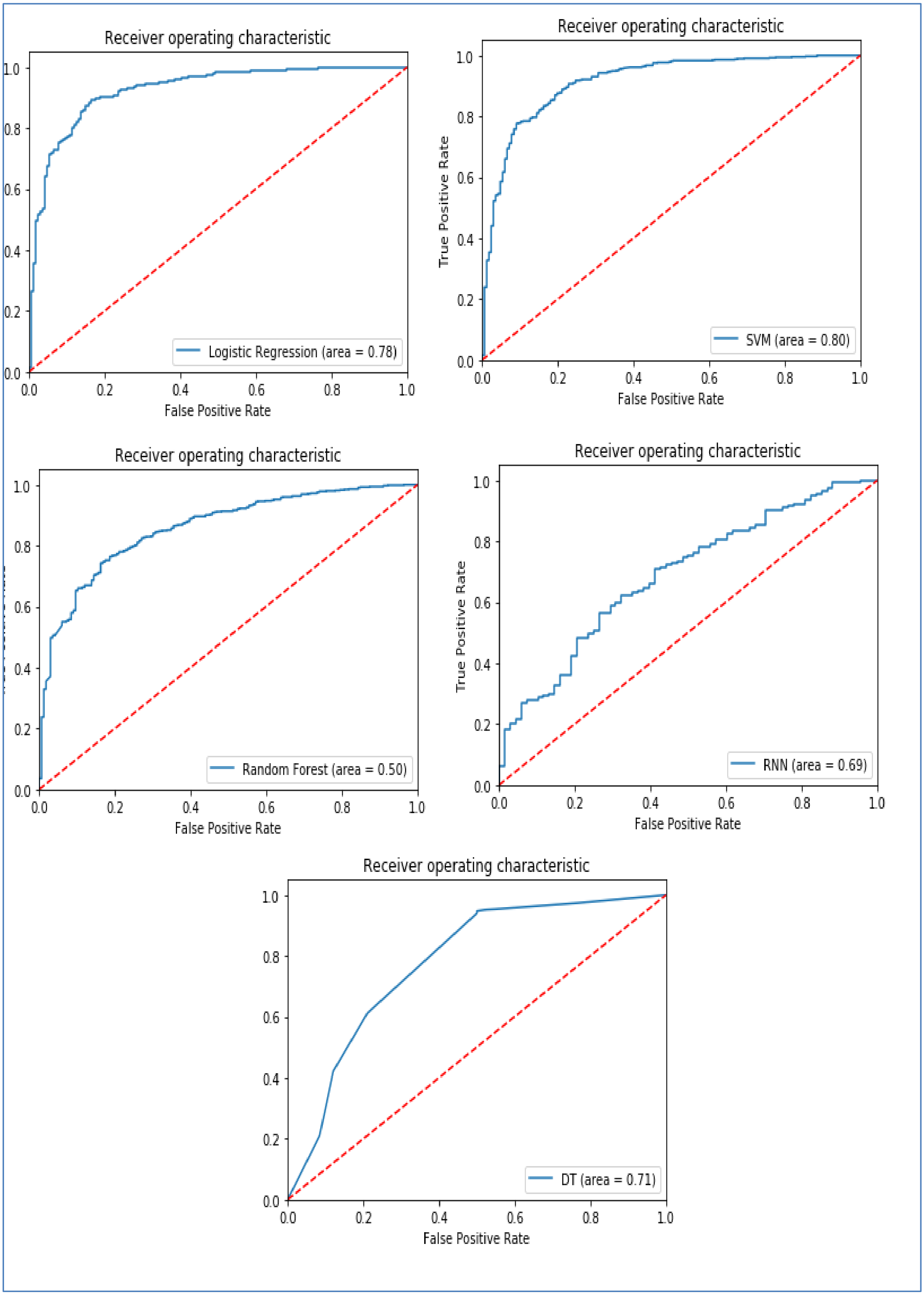

Figure 3: Receiver Operating Characteristic (ROC) Curves

6. Conclusion

In this paper, a model is proposed for Arabic sentiment classification in hotel's domain. The model consists of three layers. The first layer is pre-processing which concern with cleaning the dataset including stemming, remove stopwords and etc. The second layer is for calculating polarity and labels the text with either positive or negative using manual Arabic lexicon extracted from the same dataset. The third layer is for doing classifications using different machine learning algorithms. The experimental results show that Logistic Regression and Support Vector Machine have the highest accuracy with 89% among other classifiers. This study encourages other researchers in Arabic sentiment analysis community to improve the performance of their models and extending this study using new datasets from different domains such as books or restaurants. Also, for future works in extracting Arabic words need to have standard forms in order not miss user's opinions on the topic.

References

1. Duwairi, Rehab M., and Islam Qarqaz. "Arabic sentiment analysis using supervised classification." 2014 International Conference on Future Internet of Things and Cloud. IEEE, 2014.
2. Biltawi, Mariam, et al. "Sentiment classification techniques for Arabic language: A survey." 2016 7th International Conference on Information and Communication Systems (ICICS). IEEE, 2016.
3. El-Halees, A. 2011. Arabic Opinion Mining Using Combined Classification Approach. In Proceedings of the International Arab Conference on Information Technology, ACIT, Naif Arab University for Security Science (NAUSS), (Riyadh, Saudi Arabia), 2011.
4. Xing Fang, Justin Zhan, Sentiment analysis using product review data, Springer, Journal of Big Data (2015).
5. M. Abdul-Mageed and M. Diab, "SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis," in Proceedings of The 9th edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 2014.
6. Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of the 5th conference on language resources and evaluation (LREC'06) (pp. 417–422).
7. Ruihai Dong, Michael P. O'Mahony, and Barry Smyth, Further Experiments in Opinionated Product Recommendation, Chapter Case-Based Reasoning Research and Development, Springer International Publishing Switzerland, 2014.
8. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10 EMNLP '02 (pp. 79–86). Stroudsburg, PA, USA: Association for Computational Linguistics
9. Agarwal, Basant, and Namita Mittal. Prominent feature extraction for sentiment analysis. Springer, 2015.
10. M. Annet and G. Kondrak. A comparison of sentiment analysis techniques: Polarizing movie blogs.Advances in AI: 21st conference of the cancdian society for computational studies of intelligence, Canadian AI 2008, Windsor, Canada, May 2008, Proceedings.
11. Garg, Sonal, and Dilip Kumar Sharma. "Sentiment Classification of Context Dependent Words." Proceedings of International Conference on ICT for Sustainable Development. Springer Singapore, 2016.
12. Bhadane, Chetashri, Hardi Dalal, and Heenal Doshi. "Sentiment analysis: Measuring opinions." Procedia Computer Science 45 (2015): 808-814..

13. Xing Fang, Justin Zhan, Sentiment analysis using product review data, Springer, Journal of Big Data (2015).
14. Peleja, Filipa, and Joao Magalhaes. "Learning text patterns to detect opinion targets." Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), 2015 7th International Joint Conference on. Vol. 1. IEEE, 2015.
15. Xiaowen Ding, Bing Liu, and Philip S Yu. A holistic lexicon-based approach to opinion mining. In Proceedings of the 2008 International Conference on Web Search and Data Mining, pages 231–240. ACM, 2008.
16. Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In LREC, 2010.Xia Hu, Lei Tang, Jiliang Tang, and Huan Liu. Exploiting social relations for sentiment analysis in microblogging. In Proceedings of the sixth ACM international conference on Web search and data mining, pages 537–546. ACM, 2013.,
17. Musto, Cataldo, Giovanni Semeraro, and Marco Polignano. "A comparison of lexicon-based approaches for sentiment analysis of microblog posts." Information Filtering and Retrieval 59 (2014).
18. Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams Engineering Journal 5.4 (2014): 1093-1113. Dashtipour, Kia, et al. "Multilingual sentiment analysis: state of the art and independent comparison of techniques. " Cognitive computation 8.4 (2016): 757-771.
19. Chen, Yihua, et al. "Similarity-based classification: Concepts and algorithms." Journal of Machine Learning Research10. Mar (2009): 747-776.
20. Chin Chen Chien, Tseng You-De. Quality evaluation of product reviews using an information quality framework. Decis Support Syst 2011;50:755–68
21. Holleschovsky, Nina Isabel, and Efthymios Constantinides. "Impact of online product reviews on purchasing decisions." (2016).
22. Phillips, Paul, et al. "Understanding the impact of online reviews on hotel performance: an empirical analysis." Journal of Travel Research (2016):0047287516636481
23. Siqueira, Henrique, and Flavia Barros. "A feature extraction process for sentiment analysis of opinions on services." Proceedings of International Workshop on Web and Text Intelligence. 2010.
24. Mohammed N. Al-Kabi, Amal H. Gigieh, Izzat M. Alsmadi, Heider A. Wahsheh, Mohamad M. Haidar, Opinion Mining and Analysis for Arabic Language, (IJACSA) International Journal of Advanced Computer Science and Applications, 2014.
25. Mohammed Al-Kabi, Amal Gigieh, Izzat Alsmadi, Heider Wahsheh, Mohamad Haidar. An Opinion Analysis Tool for Colloquial and Standard Arabic. ICICS'13, April 23–25, 2013, Irbid, Jordan.
26. Alayba, Abdulaziz M., et al. "Arabic language sentiment analysis on health services." 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR). IEEE, 2017.
27. A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," Collab. Technol. Syst. (CTS), 2012 Int. Conf., pp. 546–550, 2012.
28. Hammad, Mustafa, and Mouhammd Al-awadi. "Sentiment Analysis for Arabic Reviews in Social Networks Using Machine Learning." Information Technology: New Generations. Springer International Publishing, 2016.
29. Al Shboul, Bashar, Mahmoud Al-Ayyoub, and Yaser Jararweh. "Multi-way sentiment classification of arabic reviews." Information and Communication Systems (ICICS), 2015 6th International Conference on. IEEE, 2015.
30. ElSahar, Hady, and Samhaa R. El-Beltagy. "Building large arabic multi-domain resources for sentiment analysis."

International Conference on Intelligent Text Processing and Computational Linguistics. Springer International Publishing, 2015.

31. Maher M. Itani, Lama Hamandi, Rached N. Zantout and Islam Elkabani, Classifying Sentiment in Arabic Social Networks: Naïve Search versus Naïve Bayes, 2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA).
32. Rawan T. Khasawneh , Heider A. Wahsheh , Izzat M. Alsmadi and Mohammed N. AI-Kabi, Arabic Sentiment Polarity Identification Using a Hybrid Approach, 6th International Conference on Information and Communication Systems (ICICS), 2015.
33. Mohamed Elarnaoty, Samir AbdelRahman, and Aly Fahmy, a machine learning approach for opinion holder extraction in arabic language, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012.
34. Rehab M. Duwairi and Islam Qarqaz, Arabic Sentiment Analysis using Supervised Classification, The 1st International Workshop on Social Networks Analysis, Management and Security (SNAMS - 2014), August 2014, Barcelona, Spain.
35. K. Abainia, S. Ouamour and H. Sayoud, A Novel Robust Arabic Light Stemmer , Journal of Experimental & Theoretical Artificial Intelligence (JETAI'17), Vol. 29, No. 3, 2017, pp. 557-573.
36. Catal, Cagatay, and Mehmet Nangir. "A sentiment classification model based on multiple classifiers." Applied Soft Computing 50 (2017): 135-141.
37. Keselj, Vlado. "Speech and Language Processing Daniel Jurafsky and James H. Martin (Stanford University and University of Colorado at Boulder) Pearson Prentice Hall, 2009, xxxi+ 988 pp; hardbound, ISBN 978-0-13-187321-6, $115.00." (2009): 463-466.